DOCUMENT RESUME

ED 181 407 CS 005 181

AUTHOR Latham, W.; Parry, O.
TITLE A Criterion-Referenced Test of Occupational Functional Reading Ability.
PUB DATE Sep 79
NOTE 16p.; Paper presented at the Annual Meeting of the Reading Association of Ireland (4th, Dublin, Ireland, September 13-15, 1979)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Criterion Referenced Tests; *Evaluation Methods; Foreign Countries; *Functional Reading; *Job Skills; Reading Ability; Reading Research; *Reading Skills; Research Methodology; *Test Construction; Testing; *Work Environment
IDENTIFIERS England (Sheffield); *Job Literacy

ABSTRACT

The Sheffield (Great Britain) Occupational Functional Reading Project was organized to study the functional reading required of 16-year-old school leavers who entered employment or training for a job, and to design instruments for assessing and diagnosing occupational functional reading abilities. The primary task of the project was to develop a criterion referenced test. The first steps in the project were to identify and classify job-related reading tasks, and to collect relevant reading materials. The materials were then used to construct test items, which have been stored in a computer bank for the construction of tests. Content validity of the test items was established by panels of experts. Test items were piloted and short and long forms of the prototype test were constructed. This test has been assessed for reliability and used to test a large group of school leavers prior to their actually leaving school. Followup studies of the school leavers in their workplaces are being undertaken, and their employers have been asked to rate them on job performance scales. The predictive validity of the test will be established by relating test scores to performance at work. (RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

A criterion-referenced test of occupational functional reading ability

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

W Latham and O Parry

ED181407

I Introduction

Functional reading may most simply be defined as reading related to a task or role. The study of functional reading ability, therefore, is concerned with the investigation of the different tasks in which reading is required and the assessment of both the level of ability required and the extent to which it is being achieved by the individual.

It was suggested recently, in the Green Paper: 'Education in Schools - A Consultative Document', that 'schools must prepare their pupils for the transition to adult life and work'. If this suggestion is accepted, it seems reasonable for schools to attempt to prepare their pupils for the functional reading they will encounter on leaving school, not only as an employee or trainee, but also as a citizen and consumer. Unfortunately, the schools cannot be expected to attempt such preparation as studies of functional reading ability, of the type mentioned above, have not yet been undertaken.

Such a study is the purpose of the Sheffield Occupational Functional Reading Project. In collaboration with Sheffield Metropolitan District Council the Project is studying the functional reading required of 16 year old school leavers on entering employment or training for a job, and attempting to devise assessment and diagnostic instruments of occupational functional reading ability.

II Collection of Job-related Reading Materials

As suggested above, the Project first proceeded with the identification and classification of job-related reading tasks for this group of young people. A representative sample of firms in Sheffield was visited, selected by a process previously discussed (Latham and Parry (1979)) to give a broad range of industries and types of job. Details of numbers of leavers employed, types of job. training procedures, etc. were collected, along with specimens of job-related reading materials. Industry Training Boards, Group Training Associations and Colleges of Further Education were similarly visited, where relevant. Materials were also collected from the local Careers Service.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

W. Latham
O. Parry

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CS005-181

III Classification by Type of Job

A large volume of material was thus collected, of many varying types. Reading materials included job cards, stock lists, safety regulations, descriptions of machinery, leaflets on security, personal appearance, credit card completion, and many other types. This plethora of disparate material becomes more comprehensible, however, if one groups similar types of job together. Six job-types have been used, corresponding, in the main, to those categories commonly employed by the Careers Services. These are given in Table 1.

Within each category, it has been possible to identify broad areas of overlap and many common elements. The Project is concentrating on a possible critical period in employment - the first six weeks - in which a school-leaver undergoes an adjustment to working life. Of course, technical language differs to some degree from one industry to the next, but the procedures involved in data collection resulted in a sampling of such language across the board. Here, then, we are concerned with the type of structure and content of job-related reading materials, taking vocabulary into account, and with the linguistic tasks posed by the materials encountered in the first six weeks. These factors will be considered in more detail below.

IV Test Item Construction

It is an aim of the Project to construct a criterion-referenced test of occupational functional reading ability, based on those reading materials, for use at the start of the last year at school, with those leaving at 16+. It was necessary, therefore, to construct test questions or items for this purpose.

Certain a priori criteria existed: the test must be objective (ie. have clearly identifiable correct answers for each question); it must fit into a reasonably short time span (eg. a double school-period); there must be ease of administration and marking. Further, each item must, in so far as possible, reflect a job-related reading task.

It is immediately clear that acceptable types of question are fairly restricted. There can be no essay questions, or questions requiring short written answers. Sentence completion would be a dubious type, as would cloze procedure.

Table 1: Definitions of job-types

1. Apprenticeship:	in which articles of apprenticeship are signed and agreed, with national regulations, block and/or day release training;
2. Professional:	recognised training in one of the professions, eg. articled clerk in accountancy;
3. Clerical:	general office and clerical duties which may or may not include any training;
4. Distribution Operatives/Others:	those employed in the Distribution industry; operatives: 2+ months of on-the-job training; others: no formal training;
5. Operatives/Others:	as above, all other industries;
6. Induction:	applies to reading material common to all job-types, given as part of induction into the company.

If we consider the sorts of reading material we have - lists, forms, text, etc. - it is clear that, in the main, the young person is required to find one or more pieces of data, to transfer information, to order data, etc. There is little extended writing, no sudden blanks in the middle of sentences.

Multiple choice questions suggest themselves, but, although certainly representative of a large number of job-related reading tasks, they are in themselves insufficient to represent all these reading tasks.

Three main types of item have therefore been constructed. The first is the familiar five-answer one-correct multiple choice. The second is an orally-administered type, where a number of correct answers are required for the item to be correct. The third type, known as 'action-type', requires up to four selections from up to ten possible answers. It can be seen that the first type represents tasks where a simple answer is needed. The second is for tasks where written administration may be more difficult than the actual reading material - safety signs, for example - and the third type represents those tasks where several pieces of data are required for successful completion.

V Content-validation

All materials used in the construction of items were classified according to a system adopted from the work of Sticht and his associates (Sticht et al, 1972). This classification is reproduced in Table 2.

Every item was then subject to scrutiny by panels of personnel and training officers from industry and commerce to assess its content validity. Different panels were invited, according to the type of job predominant in that industry, and asked to assess each item in terms of the representative nature of the reading passage and the relevance of the question. If the reading material failed to be representative of a type commonly encountered, the item was rejected. Similarly, should the question-part of the item be rare, irrelevant or not applicable to school-leavers in their initial period, it was either modified or rejected.

All surviving items were then examined by two linguists from the staff of the Polytechnic. Their task was to note any unseen ambiguities or unnecessary complexities and to check the content classification of each item. In fact, they performed an additional task, and suggested a classification

Table 2 Definition of content-type categories

1. Tables of content and indexes:

   Content designating the location of information with a publication.

2. Standards and specifications:

   Content setting forth specific rules or tolerance which task procedures or the completed product must conform.

3. Identification and physical description:

   Content attempting to symbolically represent an object via an identifying code (stock #, nomenclature) and/or by itemizing its distinguishing physical attributes.

4. Procedural directions:

   Content which presents a step-by-step description of how to carry out a specific job activity. Essential elements are equipment/materials/ingredients to be used, and how they are to be used, with presentation organized in a sequential step-wise fashion.

5. Procedural check points:

   Content which presents a key word or highly summarized version of what should be done in carrying out a task rather than how it should be done. This content differs from the content classified under Procedural Directions in that it assumes the user knows how to carry out the steps once reminded that the step exists and/or reminded of the decision factors which determine whether the step is required.

6. Functional description:

   Content which presents an operating (cause and effect, dependency relationships) description of some existing physical system or subsystem, or an existing administrative system or subsystem.

(By kind permission: Human Resources Research Organisation)

system based on types of linguistic task. This is given in Table 3.

VI Computerized Item-Banking

Byrne, in his comprehensive review of question-banking systems, (Byrne, 1975) suggests that "Question banks are built in order to facilitate the speedy construction of a test from a given specification" (p.6). Although the number of items surviving the content validation procedures was fairly small (about 170), both the number of data per item (31) and the complexity of the Project's requirements of such items were judged sufficient to construct a question (or "item") bank; in this case, a computerized item-banking system.

Item banks can be complex or they may be very simple. There are, however, certain irreducable minima below which a computer system becomes a waste of effort. The facilities corresponding to these minima are:

i) the selection of items on the basis of given parameters;

ii) various statistical routines;

iii) the scoring and marking of tests;

iv) the updating of the content of the item bank with new data;

v) an output of present bank contents.

These five facilities are included in the present system in varying degrees of sophistication. The item bank has been designed to be "user oriented", rather than "item-banker oriented". Virtually every user requirement has been considered and most that are feasible - and a few that are quite esoteric - have been included. Over one hundred selection criteria are available with up to 120 possible values on each; there are six different item and test analyses available; scores on tests are reported as totals and eighteen subscores; individual and group performance is reported; the provision of unreliable or inconsistent commands is not allowed and error routines pinpoint and explain each error to the user; and many other facilities.

The value of this item-banking system has been made clear to us by the ease with which it allows tests to be constructed and their results analysed. The specification of require-

Table 3: Definition of linguistic tasks

1. Referential: where the item poses a task requiring the locating of correct information;

2. Regulative: where the item requires an answer in terms of a behaviour, either an act of commission or ommission;

3. Attitudinal: where the question requires the description of a state of mind, or approach to a task;

4. Definitive: where the item requires a description of the meaning of a word or words.

ments leads, within a day, to a review of selected items and the assembly of the relevant master-copies for printing. The results for a test of 50 pupils can be available within three days. Further, the ability to investigate trends and relationships within the bank may be of great use to us in the production of diagnostic tests.

Enquiries have also been received from outside bodies as to the possibility of commissioning tests tailor-made to their requirements.

The data stored per item are given in Table 4.

VII Piloting of Items

All items surviving the content validation procedures were piloted amongst the target population. Over one hundred middle-ability Fifth Form pupils in a Sheffield comprehensive school took various sets of items in short tests. No group was smaller than seven nor larger than thirty. This initial pilot was to discover any obvious ambiguities in text or questions, to note frequently omitted items and those where a wrong answer was chosen more often than the correct answer. (See X below for item analysis details). Test administration procedures were also examined at this time.

After scrutiny, certain items were rejected or amended, these latter being repiloted.

VIII Construction of Tests

Using the item bank, piloted items were used to construct a test. The pilot study had suggested that the maximum number of items that could be answered in a double school period was about thirty. To construct a test wherein every content-type and every linguistic task was represented for each job-type would have meant the number of items necessarily exceeding thirty. Instead, a test (Form A) was constructed, with items for each job type, content-type and linguistic task. No job-type, however, had items for every content-type or linguistic task. Similarly, no content-type had items for every job.

Figure 1 indicates the format of Form A.

Hesitating, however, to reduce all the collected reading materials to only thirty-one questions, a long form of the

Table 4: Data per item in Item Bank

1. Unique, Identifying Number
2. Industrial Category (B.S.I.C.)*
3. Content Type (Sticht, 1974, Classification)
4. Job Type (Careers Service Classification)
5. Physical Format of Material (Prose, Diagrams, etc)
6. Firm, ITB, College Originally Supplying Item
7. Size of Firm
8. Item Question Type (Action, Mult.Choice, etc)
9. Answering Key
10. Order of Item on the Page
11. Code of Last Test in Which Item was Used
12. Number of the Item in Last Test
13. Number of Testees who have taken the Item
14. Number of Answer Boxes (Max. = 10)
15. Number of Testees Omitting the Item
16. Number Marking First Box
17. Number Marking Second Box
18. Number Marking Third Box
19. Number Marking Fourth Box
20. Number Marking Fifth Box
21. Number Marking Sixth Box
22. Number Marking Seventh Box
23. Number Marking Eighth Box
24. Number Marking Ninth Box
25. Number Marking Tenth Box
26. Number Correctly Responding
27. Linguistic Task Type
28. Item Number of Item replacing this one
29. Index of Discrepancies in Test-Retest
30. Discrimination Index
31. Number of Subjects in Test-Retest

* British Standard Industrial Classification, Central Statistical Office HMSO (1968)

Figure 1 Number of items by job and content type: Form A

	Apprentice	Clerical	Distribution	Op/other	Induction	Totals
Standards & specifications	1			1	3	5
Identification & Physical description	4	5		1		10
Procedural Directions	3		3			6
Procedural Checkpoints	3		1			4
Functional Description	3				3	6
Totals	14	5	4	2	6	31

Figure 2 Number of items by job and content type: Form B

	Apprentice	Clerical	Distribution	Op/other	Induction	Totals
Standards & specifications		2	4	6		12
Identification & Physical description	3	1	6	3	1	14
Procedural Directions					3	3
Procedural Checkpoints						
Functional Description		1				1
Totals	3	4	10	9	4	30

test, involving the items in Form A, together with additional items (Form B), was constructed, to give wider coverage of the job, content and linguistic areas. Thus, should a clear-cut decision not be available using Form A alone, Form B might be administered and the results added together.

Figure 2 indicates the format of Form B.

IX Test Reliability

A criterion-referenced test is one which attempts to assess an individual's ability or performance in relation to some external criterion (eg. whether or not a person can drive a car is assessed at a driving test). The reliability of such a test can therefore be seen as the reliability of classification as 'master' or 'non-master' on repeated testing. The test-retest procedure is indicated as the method of determining reliability for this type of test.

The calculation of measures of reliability remains a controversial area within criterion-referenced measurement. Various non-standard statistical techniques have been suggested - and decried - (see eg. Livingstone (1972; Harris (1972)). It was decided to use standard, product-moment correlation techniques and the decision-theoretic approach adopted by other workers in this area (Swaminathan et al 1974); Close (1977); Berk and DeGangi (1979); Subkoviak (1978)), using Cohen's K (Cohen 1960) as an unbiased estimate of the degree of agreement in classification as master/nonmaster. The coefficient was calculated for each possible score, and its maximum value taken as the score at which decisions were most reliable. This score will be used as part of the predictive validity study (see XI below).

The results for Forms A and B are given in Table 5, for repeated administration to 44 middle ability students in two schools.

It can be seen that Form A is acceptably reliable, whilst Form B is not. Further analysis of the results is being undertaken.

Table 5: Reliability Coefficient

	Pearson Correlation	Cohen's K	Optimal Cutting Score	Total no.of items
Form A	0.83	0.90	17	31
Form B	0.18	0.23	22	30

X Item Analysis

Following on from this brief discussion of problems in criterion-referenced tests, it is clear that many of the other statistics used in classical item analysis are also of limited value for the same reason: problems arising from restricted variability as between scores.

The commonest measures, difficulty and discrimination, are, in fact, of some use in criterion-referenced measurement, but not in the same sense as in classical test theory. Item difficulty is interesting but not of overwhelming importance. A test constructor may require a mixture of easy and hard items, but more often in criterion-referenced measurement he requires a test of a certain area. If the test has content validity, the difficulty of the items is secondary. Discrimination indices for an item are not valuable in terms of seeing how the item contributes to spreading out scores on the test, but to see whether an item positively or negatively discriminates. The latter is as much an indicator of a poor item in criterion-referenced measurement as in norm-referenced measurement (eg. Smith, 1974).

XI Predictive Validation

The ultimate value of a test of functional skill is its ability to predict adequate performances. In this case, the test of occupational functional reading ability should predict performance on job-related reading tasks in the first six weeks of employment or training for a job.

As an initial measure of this predictive validity, about 150 fifth-form school-leavers, mainly of average ability, were given Form A of the test, and the Edinburgh Reading Test, Form 4. This latter is a norm-referenced test for comparison purposes; extrapolated quotients were used for those pupils already 16 years old.

These pupils have now left school and have found or are seeking employment. The Sheffield Careers Service has kindly allowed access to its files, so that each subject may be followed up when in employment. Each employer is being asked to allow the young person's immediate supervisor to fill in a rating form, assessing relevant aspects of job-performance on tasks involving and not involving reading. These include perseverence, need for oral guidance, speed, etc.

Test scores and ratings will then be related via multiple and linear regression analyses, to establish the test's predictive validity.

XII Summary

The Sheffield Occupational Functional Reading Project has as its remit the development of a criterion-referenced test of occupational functional reading ability of pupils, leaving at 16+ years of age, in their last year at Sheffield secondary schools. Related diagnostic tests are also to be constructed.

Progress towards these aims has taken the form of identification and classification of job-related reading tasks and collection of relevant reading materials. These have been used to construct test items, the data about which have been banked by computer. Content validity was established by the use of panels of experts. Test items were piloted and a short and long form of a prototype test constructed. This test has been assessed for reliability and then used to test a large group of school-leavers, prior to their leaving school. These testees are now being followed up at work and their employers asked to rate them on job-performance scales. The predictive validity of the test will be established by relating test score to performance at work.

References

Berk R A & De Gangi G A — Technical Considerations in the Evaluation of Pediatric Motor Scales American Journal of Occupational Therapy 1979, 33, No.4 pp.240-244

Byrne C J — Computerized Question Banking Systems Institute of Educational Technology, The Open University, Milton Keynes 1975

Close J S — Personal communication

Cohen J — A coefficient of agreement for nominal scales Educational & Psychological Measurement, 1960 20, 37-46

Department of Education and Science — Education in Schools - A Consultative Document HMSO Cmnd.6869 1977

Harris C W — An interpretation of Livingstone's reliability co-efficient for criterion-referenced tests. Journal of Educational Measurement, 1972 9, 27-29

Latham W & Parry O — Functional Reading and the Schools in Growth in Reading ed. Thackray D V London, Ward Lock Educational, 1979

Livingstone S A — Criterion-referenced applications of classical test theory Journal of Educational Measurement 1972, 9, 13-25

Smith C W — Criterion-referenced Assessment in Contemporary Issues in Educational Testing ed. Crombag, H F & De Gruyter, D N The Hague, Mouton, 1974, pp 138-147

Sticht T G, Caylor J S, Kern R P and Fox L C — Project REALISTIC: determination of adult functional literacy skill levels. Reading Research Quarterly, 1972, 7, 424-465

Subkoviak M J — Empirical Investigation of Procedures for Estimating Reliability for Mastery Tests Journal of Educational Measurement 1978, 15, No.2, pp 111-116

Swaminathan H, Hambleton R K & Algina J — Reliability of Criterion-Referenced Tests: a decision-theoretic formulation Journal of Educational Measurement, 1974, 11, 263-7